An Introduction to Data Quality Management and Data Quality Auditing

With thanks to Khulisa Management Services

Content of Workshop Session

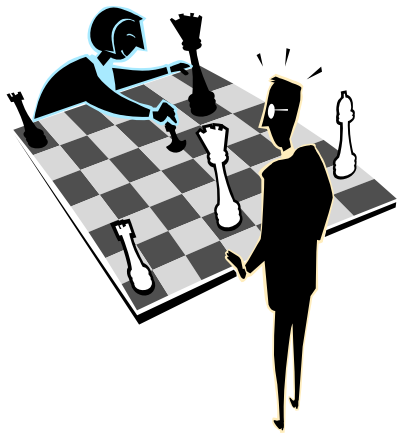
- What is Data Quality?
- The criteria for data quality
- Data quality auditing
- Constructing a data quality plan

Why is data quality important?

- Program decision making
- Sharing program information
- Reporting/Accountability

What is Data Quality ?

- Chess game of cost versus quality
- Criterion based evaluation of data
- Criterion based system of data management



Data Quality Criteria

- Validity
- Reliability
- Timeliness
- Precision
- Integrity

Definition of Validity

- A characteristic of measurement that a tool actually measures what the researcher actually wants to measure
- Have we actually measured what we intended?

Threats to Validity

- Definitional issues
- Proxy measures
- Inclusions / Exclusions
- Data sources

Validity: Questions to ask yourself…

- Is there a relationship between the activity or program and what you are measuring?
- What is the data transcription process? Is there potential for error?
- Are steps being taken to limit transcription error (e.g., double keying of data for large surveys, built in validation checks, random checks)?

Validity: Questions to ask yourself…

- If there are data errors, what do you do with that information?
- If raw data need to be manipulated, are the correct formula being applied and applied consistently (e.g. from site to site, over time)?
- What do I do if I have a missing/incomplete data set?
- Are final numbers reported accurately (e.g. does the total add up)?

Definition of Reliability

- ‘A characteristic of measurement concerned with consistency’
- Can we consistently measure what we intended?

Threats to Reliability I

Persona

Time

Place

Threats to Reliability II

- Collection methodologies
- Collection instruments
- Sampling frameworks
- Personnel issues
- Analysis and manipulation methodologies

Reliability: Questions to Ask Yourself…

- Is the same instrument used from year to year, site to site?
- Is the same data collection process used from year to year, site to site?
- Are there procedures in place to ensure that data are free of significant error and that bias is not introduced (e.g., instructions, indicator information sheets, training, etc.)?

Definition of Timeliness

- The relationship between the time of collection, collation and reporting to the relevance of the data for decision making processes.
- Does the data still have relevance and value when reported?

Threats to Timeliness

- Collection frequencies
- Reporting frequencies
- Time dependency

Timeliness: Questions to ask yourself…

- Are data available on a frequent enough basis to inform program management decisions?
- Is a regularized schedule of data collection in place to meet program management needs?
- Are data from within the policy period of interest (i.e. are the data from a point in time after the intervention has begun)?
- Are the data reported as soon as possible after collection?

Definition of Precision

- Accuracy (measure of bias)
- Precision (measure of error)
- Is the margin of error in the data less than the expected change the project was designed to effect?

Threats to Precision

- Source error / bias
- Instrumentation error
- Sampling error
- Transcription error
- Manipulation error

Precision: Questions to ask yourself…

- Is the margin of error less than expected change being measured?
- Are the margins of error acceptable for program decision making?
- Have issues around precision been reported?
- Would an increase in the degree of accuracy be more costly than the increased value of the information?

Definition of Integrity

- Measure of ‘truthfulness’ of the data
- Is the data free from ‘untruth’ introduced by either human or technical means, willfully or unconsciously?

Threats to Integrity I

Time

Temptation

Technology

Threats to Integrity II

- Corruption, intentional or unintentional
- Personal manipulations
- Technological failures
- Lack of audit verification and validation

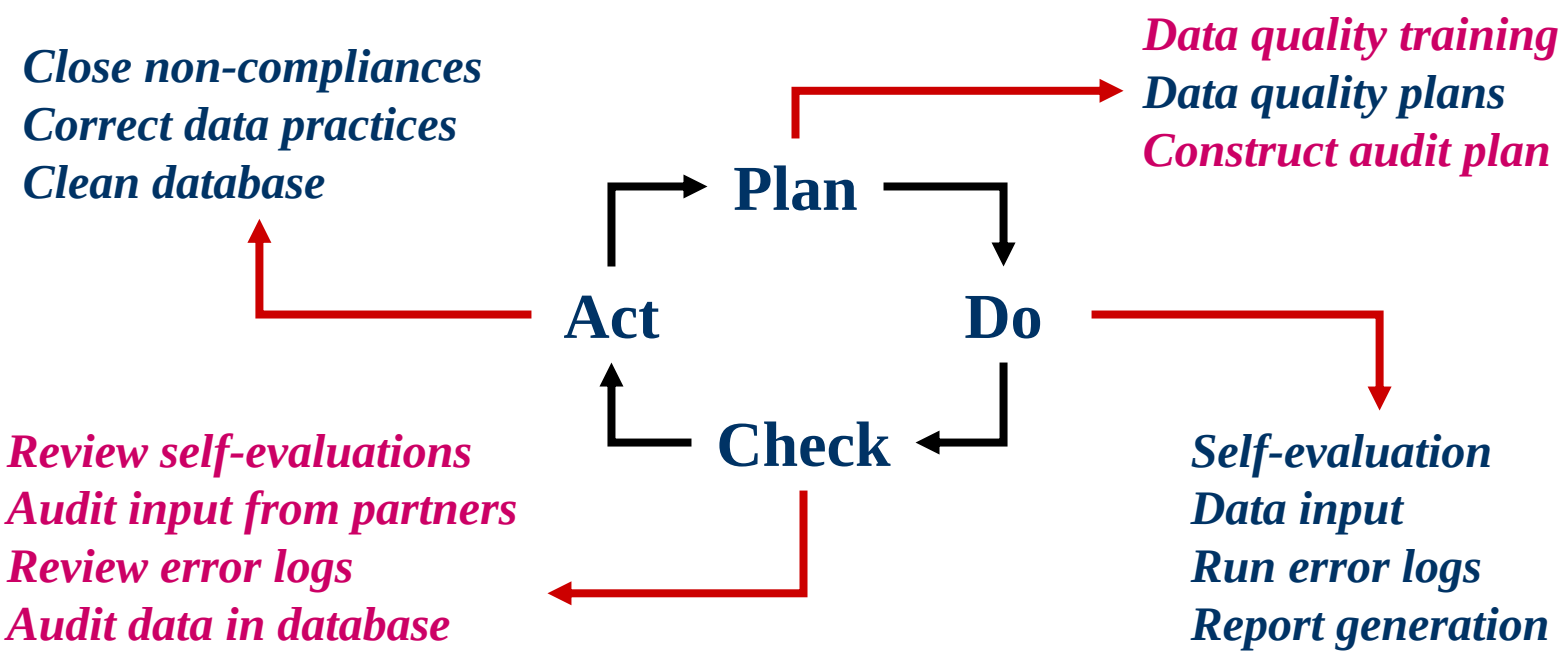
Integrity: Questions to ask yourself…

- Are there risks to data are manipulated for personal or political reasons?
- What systems are in place to minimize such risks?
- Has there been an independent review?

Data Quality Audits

- Phase 1: Self-assessment and Process Assessment
- Phase 2: Verification and Validation
- Internal audit
- External audit

DQA Process



The auditor is responsible for the areas indicated in red

Findings from Data Quality Audits

- “One-man show”
- Putting all indicators on one Indicator Information Sheet
- Definitions of Indicators – Number of people counseled and tested (indirect)
- Every partner had simple calculation errors which could be corrected if people had someone double checking data entry/manipulation

Findings from Data Quality Audits

- Clear data collection tools and instructions
- Version Control
- Missing data
- Audit Trail (data warehouse; feedback/instructions)
- Risks involved in using volunteers

Findings from Data Quality Audits

- Version Control
- Missing data
- Audit Trail (data warehouse; feedback/instructions)
- Risks involved in using volunteers

Findings from Data Quality Audits

- Issues around double-counting
- Not collecting data routinely so ‘reporting flurry’
- Integrity – incentives for over-reporting
- Documentation of what was reported
- Effect of staff turnover
- Involving staff in M&E – definitions of indicators, value of data, data use

The Data Quality Plan

- Operational Plan for managing data quality
- Related to the Indicator Information Sheets
- Includes a Data Quality Risk Analysis
- Includes an audit trail reference

Establishing Data Quality Risks I

Overall Effect on Data Quality	Probability of Error Occurring			
	(4) - Constantly	*(3) - Frequently*	*(2) - Occasionally*	*(1) - Unlikely*
(4) - Catastrophic	**16**	**12**	**8**	**4**
(3) – Critical	**12**	**9**	**6**	**3**
(2) - Marginal	**8**	**6**	**4**	**2**
(1) - Negligible	**4**	**3**	**2**	**1**

Establishing Data Quality Risks II

Risk Score	Risk Type	Remedial Action
9 - 16	**High Risk**	Establish contingency plan to reduce risk, verify and validate *prior to each reporting episode*, maintain strict audit trail.
4 - 8	**Medium Risk**	Establish contingency plan to reduce risk, verify and validate *prior to annual return*, maintain strict audit trail.
1 - 3	**Low Risk**	No immediate action required; risk could be managed through normal internal audit processes.

Minimizing Data Quality Risks

- Technology
- Competence of personnel
- Documentation and audit trails
- Outsourcing

Exercise

- Complete a Data Quality Assessment for one of your indicators
- What are the data quality risks associated with the indicator?
- Draft a Data Quality Plan for the Indicator – answers the questions – what are the risks and what is your plan for managing them?

Thank you !